Representing Dialog Progression for Dynamic State Assessment *

Sophie Rosset, Lori Lamel
LIMSI-CNRS, BP 133
91403 Orsay cedex, France
{rosset, lamel}@limsi.fr

Abstract

While developing several spoken language dialog systems for information retrieval tasks, we found that representing the dialog progression along an axis was useful to facilitate dynamic evaluation of the dialog state. Since dialogs evolve from the first exchange until the end, it is of interest to assess whether an ongoing dialog is running smoothly or if is encountering problems. We consider that the dialog progression can be represented on two axes: a Progression axis, which represents the "good" progression of the dialog and an Accidental axis, which represents the accidents that occur, corresponding to misunderstandings between the system and the user. The time (in number of turns) used by the system to repair the accident is represented by the Residual Error, which is incremented when an accident occurs and is decremented when the dialog progresses. This error reflects the difference between a perfect (i.e., theoretical) dialog (e.g. without errors, miscommunication...) and the real ongoing dialog. One particularly interesting use of the dialog axis progression annotation is to extract problematic dialogs from large data collections for analysis, with the aim of improving the dialog system (Wright-Hastie, Prasad and Walker, 2002). In the context of the IST Amities project (Amities, 2001-2004) we have extended this representation to the annotation of human-human dialogs recorded at a stock exchange call center and intend to use the extended representation in an automated natural dialog system for call routing.

1 Introduction

At LIMSI we have experience in developing several spoken language dialog systems for information retrieval tasks (Bonneau-Maynard et al., 1993; Gauvain et al., 1997; Lamel et al., 2000; Shao et al., 1998). In our view, spoken language systems should provide a natural, user-friendly interface with the computer, allowing easy access to the stored information. We have developed applications in two classes of SLDSs, telephone-based and kiosk-based. Telephone based services are a natural area for spoken dialog systems as the only means of interaction with the machine are via voice and have thus been the focus of many development efforts. Our activities in this area have been mainly in the context of European projects and a French language action launched by the AUPELF-UREF (Bonneau-Maynard and Devillers, 2000).

As more natural SLDSs are developed it is becoming apparent that the dialog manager is a crucial aspect of the system, and design decisions and functionality influence all other system components (Rosset, Bennacef and Lamel, 1999). Some considerations concern strategies for error

*This work was partially financed by the European Commission under the IST-2000-25033 AMITIES project.

detection and correction, and conflict resolution. One of the most important problems is to endow the automatic dialog system with the capability of detecting and repairing problems which arise during conversation. If the system is able to detect that it is encountering problems, it can adapt its communication strategy or transfer the call to an human operator. However, systems are rarely able to determine the cause of the problem. One way to adapt the conversation's strategy when problems occurs is to adapt the reaction to the gravity of the problem. Systems generally consider that an error has occurred when there is a contradiction between two (or more) acquired items.

In this paper, we present a representation of dialog progression which can be used by the dialog manager to dynamically adapt its dialog strategy to the state of the dialog, in order to achieve high dialog success rates. This representation has been used to automatically annotate a large corpus of human-machine dialogs for an information retrieval task (Lamel et al., 2000). The dialog axes representation for use in automatic dialog systems is discussed in Section 2. In Section 3 we explore the use of dialog progression annotation for human-human dialogs. Our initial study quickly led to the realization that the representation was insufficient for the uncontrolled nature of the call center human-human dialogs being analyzed in the IST-AMITIES project (Amities, 2001-2004). An extended annotation for these dialogs is proposed.

2 Axes for Progression in Spoken Language Dialog Systems

In this section we overview the characteristics of an automatic spoken language dialog system (SLDS), and described the dialog axes progression annotation scheme.

2.1 Core Technologies for SLDS

The main components of a spoken language dialog system are a speech recognizer, a natural language analyzer, and a dialog manager, which controls the information retrieval component including database access and response generation, and a speech synthesizer. The dialog manager is the controller of the overall system as it manages contextual understanding, the dialog history, information retrieval and response generation. The generation component outputs a natural language response based on the dialog state, the caller's query, and the information returned after database access. As more natural SLDSs are developed it is becoming apparent that the dialog manager is a crucial aspect of the system, and design decisions and functionality influence all other system components (Rosset, Bennacef and Lamel, 1999). Some considerations concern strategies for error detection and correction, and conflict resolution.

2.2 Dialog Manager and Dialog Progression Representation

To support a user-friendly, mixed initiative dialog, the system should support negotiation, navigation (that is detection of topic or task changes), and to the extent possible, be able to detect and deal with errors. When the dialog is going well, the user should be able to express him/herself freely, providing information in any order. If the dialog is not progressing, the system should guide the user. Long dialogs are often a good indication that the user is experiencing problems. Therefore, we try to minimize the number of dialog turns, in order to rapidly aide the user to obtain their desired information. To support different user needs, a two-level dialog strategy has been implemented, in which a mixed-initiative dialog is combined with a system-directed dialog in case a problem is detected in obtaining important information. When the second level, or constrained dialog is active, the speech recognizer makes use of a dialog-state dependent language model.

All dialogs evolve, from the first exchange until the end. In order to achieve high successful dialog completion rates, it can be interesting to assess whether an ongoing dialog is running smoothly or if is encountering problems. We consider that the dialog progression can be represented on two axes: a Progression axis, which represents the "good" progression of the dialog and an Accidental axis, which represents the accidents that occur between the system and the user. These axes are represented by respective values, P and A. At each turn, one of the two axes is incremented by 1 (P when all is ok and A when an accident has occured). The number of turns (T) in the dialog is equal to the sum A+P. Figure 1 shows the evolu-

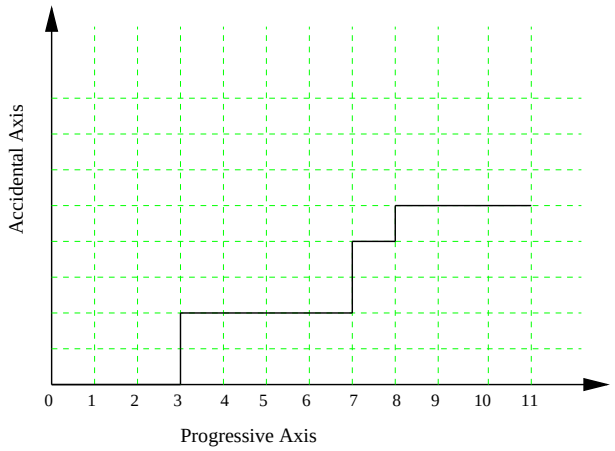

Figure 1: Evolution of a dialog according to the Progression (P) and Accident (A) axes.

tion of the course of a dialog on the two axes.

The higher the value of A, the more the dialog encounters problems. Similarly, the longer it takes to return to the progressive axis, the more the dialog degrades. A third value, the Residual Error (RE) is used to represent the time (in number of turns) used by the system to repair the accident. The residual error, which is incremented when the A value is incremented and decremented when the P value is incremented, represents the difference between a perfect, theoretical dialog (e.g. without errors, miscommunication...) and the real dialog. This residual error makes easy to know if at a particular moment T, how close the ongoing dialog is to a certain standard for a reasonable dialog.

Since the dialog duration as function of task completion seems to be an important factor for user satisfaction (see (Walker, Boland and Kamm, 1999) for example), we added two other values: the maximum theoretical duration in turns (T_{max}) and δT. In order to dynamically adapt to the dialog flow, T_{max} is updated according to the model of the task for each interpreted statement. For example, in the case of a train travel information task, a certain number of elements are needed for database access. Therefore, T_{max} is set to the number of elements that are required, assuming that in an error-free dialog at most one exchange is needed per item. δT is the the difference between the number of exchanges T and T_{max}.

The dialog duration has a relative importance according to the preceding dialog. If T_{max} is exceeded following an error (or an augmentation of the A axis) then the error is more significant and detrimental for the continuation of the course of the dialog than if T_{max} being exceeded does not follow an error. In the former case the long dialog duration may indicate that the systems is having a hard time acquiring the items needed for database access, whereas in the latter the system may simply be guiding the user. Some users prefer to be guided by the system and are unperturbed by a long dialog as long as there are not recurrent errors. The T_{max} and δT values can be applied to the whole dialog and to different sub-tasks. The goodness of the dialog can be assessed according to a score:

$$Score = \frac{-(\beta A - P + \alpha \delta T)}{\beta A + P - \alpha \delta T},$$

where $0 \leq \alpha < 1$ et $0 \leq \beta \leq 1$ are weights for A and δT ($\delta T = (T - T_{max})$

Figure 2 shows an example of how these values are used. In this example, the score is calculated both globally (S_g) and locally (S_l). For the global score, the T_{max} is set to 8. For the local score, the T_{max} is 3. After the fifth turn, the score indicates to the system that problems have occurred, even though there was not an error. That can be the sign of a novice user.

This kind of annotation was implemented with deductive rules in the Arise system. We are currently working on adapting it for use in the IST-AMITIES project. Figure 3, gives the score progression for two sample dialogs taken from the ARISE corpus. The dialog on the left proceeds without any major problems, whereas the dialog on the right is quite problematic given the low scores and long duration.

3 Axes for Annotation of Human-Human Dialogs

In the IST-AMITIES project (Amities, 2001-2004) we are analyzing large corpora of human-human (and eventually human-computer) dialogs in order to develop empirically-based natural spoken dialog systems. The corpus used in this study

S: Quels renseignements désirez-vous? *(What information do you want?)* **[0 0 0]** S_g = 1, S_l = 1

U: je voudrais les horaires des trains *(train timetable information)*
R: {fw} les horaire des train
S: Oui, de quelle ville partez-vous? *(what is your departure city?)* **[1 0 0]** S_g = 1, S_l = 1

U: Albertville
R: Albertville
S: Bon, vous partez de Albertville, quelle est votre destination? *(from Albertville, what is your destination?)* **[2 0 0]** S_g = 1, S_l = 1

U: Dijon
R: Dijon
S: Alors, de Albertville à Dijon, quel jour partez-vous ? *(from Albertville to Dijon, what date are you traveling?)* **[3 0 0]** S_g = 1, S_l = 1

U: je veux partir le [bb] le *(I want to leave the [bb] the)*
R: je-veux partir le simple
S: Veuillez indiquer la date de votre départ, par example le 4 juin. *(please indicate your departure date, for example June 4th)* **[3 1 1]** S_g = 0.787234, S_l = 0.6875

U: le cinq cinq juin *(the 5th 5th of June)*
R: le cinq juin
S: vendredi 5/6/98 de Albertville à Dijon à quel moment de la journée voulez-vous partir? *(Friday 5/6/98 from Albertville to Dijon, what time do you want to leave?)* **[4 1 0]** S_g = 0.814815, S_l = 0.74359 $T_{sup} = T - T_{max}$ = 5 - 8 = -3;/ $T - T_{max}$ = 5 - 3 = 2

Figure 2: Example dynamic dialog evaluation use the dialog axes. **S:** system prompt; **R:** automatic transcription; **U:** orthographic transcription. **[P,A,RE]**: the three parameters for progression, accident and residual error. S_g = global score and S_l = local score; $\beta = 0.5$ et $\alpha = 0.3$ reflect the dialog state which generates the given system prompt.

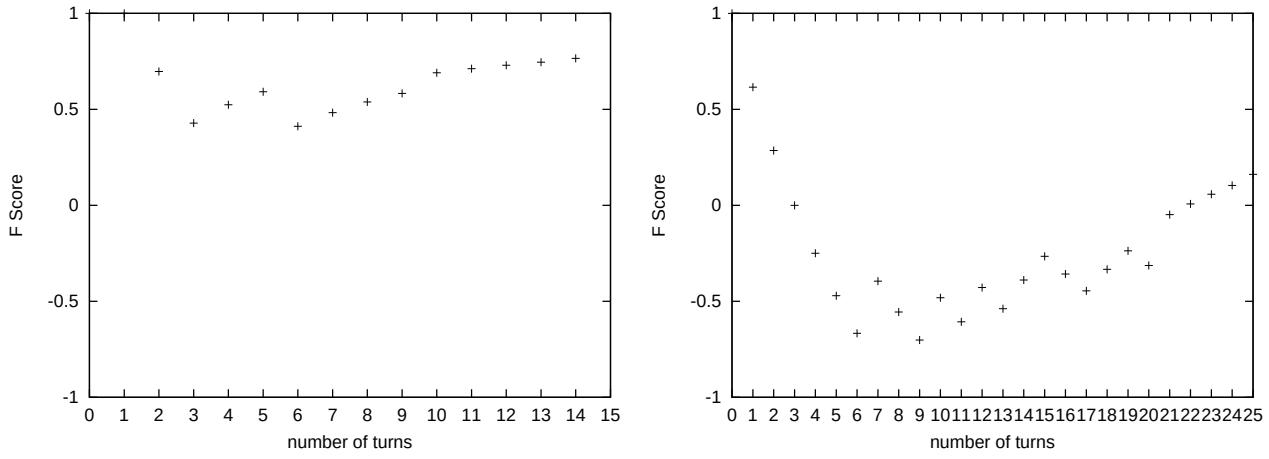

Figure 3: Score progression for two sample dialogs from the ARISE corpus. The dialog on the left proceeds reasonably well, whereas the dialog on the right is problematic.

is described, and extensions to the dialog progression axes annotation are proposed.

3.1 Corpus

In this study we make use of a set of real agent-client dialogs recorded at a Web-based Stock Exchange Customer Service center. These recordings were made for purposes independent of this study, and have been made available for use in developing an automated call routing service within the context of the AMITIES project. The service center can be reached via an Internet connection or by directly calling an agent. While many of the calls are concerned with problems in using the Web to carry out transactions (general information, complicated requests, transactions, confirmations, connection failures), some of the callers simply seem to prefer interacting with a human agent. A corpus of 100 agent-client dialogs (from 4 different agents) in French has been orthographically transcribed and annotated. The dialogs cover a range of investment related topics such as information requests (services, commission fees, stock quotations), orders (buy, sell, status), account management (open, close, transfer, credit, debit) and Web questions and problems. In the 100 annotated dialogs, there are 5229 speaker turns after excluding overlapping speech segments (overlapping speech is known to be a frequent phenomenon in conversations). The corpus contains a total of 44.1k words, of which 3k are distinct. The average dialog length is 50 utterances (min 5, max 227), the average sentence length is 9 words (min 1, max 128).

3.2 Axes

In attempting to annotate an initial set of agent-client dialogs it became apparent that the annotation scheme used for human-machine dialogs needed to be extended in order to better cover the more varied human-human dialogs. In the human-human dialogs we encountered some speech turns which are not directly concerned with the task, and therefore do not match either of the labels P or A. Moreover, some phenomena, like backchannel acknowledgments, which are helpful for the communication management do not directly contribute to the progression of the dialog with an A or a P value, insofar as the task is concerned. Thus, we decided to add two new values to label these types of utterances: Out-of-Task and Backchannel. The P, A and RE values remain unchanged (i.e., they are not incremented or decremented) when the turn is not directly concerned with the task or when it is only a backchannel (except when the backchannel marks a repair). Thus 5 dialog axes values are used to annotate the 100 agent-client dialog corpus: T (turn of speech), P (progression), A (accident), RE (residual error) and OT (out of task). The annotation is marked only on the agent's turn. The entire corpus contains 1136 progressions. There are a total of 252 accidents, which occured in 70 dialogs. 35 of these dialogs have unrepaired accidents at the end of the dialog, with a summed total of 88 residual errors across all dialogs.

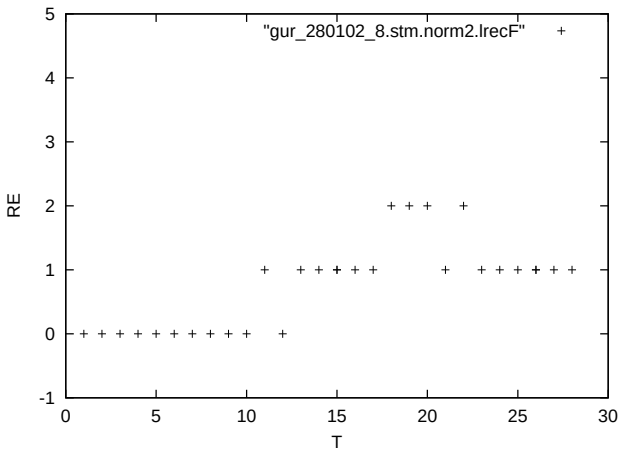

Figure 5: Residual error vs. dialog turn in an Agent-Client dialog.

Figure 4 shows an excerpt from one of the human-human dialogs along with the progression axis labels. The full dialog is much longer. Figure 5 plots the Residual Error as a function of the dialog turn for this excerpt, giving a graphical representation of the dialog progression. It can be noticed that several turns may be needed to get a dialog back on track and that not all errors are corrected.

4 Conclusions and Perspectives

In previous work developing spoken language dialog systems, we found that automatic annotation of the dialog progression according to 3 axes (for progression, accidents, and residual error) was

A: d'accord *(ok)* **[9 2 0 0 1]**

C: bon et euh ça c'est une première chose deuxième chose je j'ai fait des opérations le vingt trois et le vingt cinq *(that's the first thing second thing I did some transactions the 23rd and 25th)*
A: vingt trois et vingt cinq *(23 and 25)* **[10 2 0 0 1]**

C: oui euh et euh euh *(yes uh uh)*
A: hein *(huh)* **[11 2 1 1 1]**

C: oui oui et les c'est pour ça que j'ai attendu votre appel je préférais plutôt que d'en discuter et euh les les taux pratiqués ne sont pas du tout ceux euh qui sont pratiqués habituellement hein donc il y a il y a des erreurs euh *(yes yes and that's why I was waiting for your call I would prefer rather than to discuss uh the the rates used are not those generally used uhm therefore there are there are some errors uh)*
A: qui a été prélevé peut-être non *(which was charged maybe or not)* **[12 3 1 0 1]**

C: oh je ne je ne sais pas on puis il y a des moments où il y a eu encore des problèmes informatiques euh donc euh or que j'avais eu à ce moment-là m'a dit que euh il allait resignaler où vous aviez changé de ... *(oh I I don't know and sometimes there are some technical problems uh uh if I knew at this time ...)*
A: c'est deux opérations *(it is 2 transactions)* **[13 3 2 1 1]**

C: plusieurs hein il y en a plusieurs j'en ai fait quatre ou cinq à peu le mercredi vingt trois et le vendredi vingt cinq *(several uh there are several that I did 4 or 5 Wednesday the 23rd and Friday the 25th)*
A: ok donc ca je le note *(ok I'm writing them down)* **[14 4 2 1 1]**

Figure 4: Example of dialog progression annotation with the 5 axe values **[T,P,A,RE,OT]** and **[T]**: Turn number, **[P]**: Progression, **[A]**: Accident, **[RE]**: Residual Error, and **[OT]**: out-of-task. The excerpt corresponds to turns 9 through 14 in Figure 5. One uncorrected residual error remains at turn 14.

useful for dynamic assessment of dialog quality. These axes were used by the system to automatically evaluate (by itself) the progression of the ongoing dialog, and to use this to modify the dialog strategy when problems are suspected.

When the same annotation scheme was applied to more complicated human-human dialogs, it was found that the labels needed to be extended to cover previously unobserved events (out-of-task sentences and backchannel). Our impression is that this annotation scheme does not seem to be very useful for annotating very free human-human dialogs such as the agent-client dialogs recorded at the Web-based Stock Exchange Customer Service center. We believe that the scheme could be successfully applied to more constrained human-human dialogs such as those at financial call centers where a strict protocol is followed in order to obtain access to a variety of services.

In the context of the Amities project this annotation scheme will be used to label a set of directed human-human dialogs collected at financial service call centers. These dialogs will also serve for the development of the dialog manager of an automatic system for call routing in French and English. Since this annotation scheme can be done (*a posteriori*) with or without comparison to the true transcription or with the output from speech recognizer, it can also be used to extract problematic dialogs from large data collections for further analysis in order to improve the dialog system (Wright-Hastie, Prasad and Walker, 2002).

References

AMITIES, http://www.dcs.shef.ac.uk/nlp/amities

1st SIGDIAL workshop on Discourse and Dialog, http://www.sigdial.org/sigdialworkshop/program.html

J. Allen and M. Core, "Draft of DAMSL: Dialog Act Markup in Several Layers," October 1997. http://www.cs.rochester.edu/research/trains/annotation

H. Bonneau-Maynard, L. Devillers, "A Framework for Evaluating Contextual Understanding," ICSLP'00.

H. Bonneau-Maynard, J.L. Gauvain, D. Goodine, L.F. Lamel, J. Polifroni, S. Seneff, "A French Version of the MIT-ATIS System: Portability Issues," *Eurospeech'93*.

J.L. Gauvain, S.K. Bennacef, L. Devillers, L.F. Lamel, and S. Rosset, "Spoken Language component of the MASK Kiosk" in *Human Comfort & Security of Information Systems*, K.Varghese, S.Pfleger (Eds.), Springer-Verlag, 1997.

L. Lamel S. Rosset, J.L. Gauvain, S. Bennacef, M. Garnier-Rizet, B. Prouts, "The LIMSI ARISE System," *Speech Communication*, **31**(4), pp. 339-353, Aug. 2000.

D. Luzzati, "Recherches sur le dialogue homme-machine: modèles linguistiques et traitements automatiques", Thèse d'état, Paris III, 1989.

S. Rosset, S.K. Bennacef, L.F. Lamel, "Design Strategies for Spoken Language Dialog Systems," *Eurospeech'99*.

JR. Searle, D. Vanderveken, "Fondations of Illocutioanry Logic," Cambridge: CUP, 1985.

J. Shao, N.E. Tazine, L. Lamel, B. Prouts, and S. Schröter, "An Open System Architecture for Multimedia and Multimodal User Interface," *3rd TIDE Congress*, Helsinki, June, 1998.

M. Walker, J. Boland, C. Kamm, "The utility of elapsed time as a usability metric for spoken dialogue systems", ASRU'99.

H. Wright-Hastie, R. Prasad and M.A. Walker, "What's the Trouble: Automatically Identifying Problematic Dialogs in DARPA Communicator Dialog Systems," *Proc. ACL'02*.